Connected Devices Classification using Feature Selection with Machine Learning

Fatima Zahra Fagroud, Hicham Toumi, El Habib Ben Lahmar, Khadija Achtaich, Sanaa El Filali, and Youssef Baddi

***Abstract*— In the latest years, with the integration of not only traditional networking devices but also a wide variety of Internet of Things (IoT) devices, connected devices retrieval gain much importance. This retrieval can be carried out through the use of various information research tools such as search engines. The devices type represents one of the most important attributes of connected devices retrieval, which helps effectively to give accurate results during the retrieval process. Efficient and sample connected devices classification and type identification represent a highly difficult task due to the wide variety of existing and evolving devices type. Identifying appropriate characteristics of connected devices may help simplify the classification process, which requires the integration of feature selection methods. For this, we analyze in this work the effect of feature selection to achieve a high performance and high accuracy of connected devices classification. To this end, we use different feature selection methods and evaluate these methods by applying a set of machine learning models. To extract the most representative features of our datasets we employed univariate feature selection, Recursive Feature Elimination (RFE), and Tree-based feature selection (Random Forest). XGBoost, Decision Tree, and Random Forest are applied for performance evaluation based on the extracted feature. The evaluation results show that the selection of important features helps to improve the accuracy of connected devices classification using machine learning classifiers.**

***Index Terms*—Connected Devices, Feature Selection, Classification, Internet of Things, IoT**

I. Introduction

IoT is a dynamic global network infrastructure that seamlessly integrates physical and virtual "things" that have identities, physical attributes, and virtual personalities into the information network. IoT is self-configuring capabilities using standard and interoperable communication protocols [1]. It represents a network of connected things which are connected to the Internet and other things, and able to gather and share information related to the way they are employed and almost the environment around them. [2]

Since the appearance of IoT solutions, IoT solutions suppliers keep proposing more and more inventive connected objects. An important number of these solutions are present in our daily lives to respond to our needs, help us and realize a set of tasks. This wide variety and the exponential growth of connected devices (According to a study of GSMA the number of IoT connections in the world attained 25,2 billion connections and the global market of IoT rise to 1100 billion dollars excluding hardware (devices, modules, and chips) by 2025 [3]) introduce a new challenge in the IoT research topic, which is device classification and type identification.

Knowing the type of device connected to the network will help to enforce security and greatly manage the network. For example, knowing that a device is a security camera from a specific manufacturer can help the network administrator to specify filtering rules that will not allow the camera to do anything else than what it is expected to do. Device type recognition can also be used to block access to the network of devices considered to be vulnerable. IoT device recognition can also be used by an attacker to discover vulnerable IoT devices by performing passive network traffic analysis. [4]

Connected devices classification and type identification is a challenging subject in IoT topic which has gained great importance recently by researchers. A good deal of research has been conducted on connected devices classification and type identification using machine learning (ML) techniques which are based on network traffic data in most studies, but little of them use feature selection methods to select the better and most important features. In this work, we study the effectiveness of using feature selection methods to provide better classification performance. We applied some feature selection methods on data collected from shodan and then applied machine learning models on the features selected.

Manuscript received September 16, 2021; revised March 10, 2022.

F. Z. Fagroud is a PhD candidate at Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, BP 7955 Sidi Othman Casablanca, Morocco (e-mail: fagroudfatimazahra0512@gmail.com).

H. Toumi is a Professor at Higher School of Technology - Sidi Bennour Chouaïb Doukkali University El Jadida, Morocco (e-mail: toumi.h@ucd.ac.ma).

E. Ben Lahmar is a Professor at Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, BP 7955 Sidi Othman Casablanca, Morocco (e-mail: h.benlahmer@gmail.com).

K. Achtaich is a Professor at Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, BP 7955 Sidi Othman Casablanca, Morocco (e-mail: k.achtaich@gmail.com).

S. El Filali is a Professor at Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, BP 7955 Sidi Othman Casablanca, Morocco (e-mail: elfilali.sanaa@gmail.com).

Y. Baddi is a Professor at Higher School of Technology - Sidi Bennour Chouaïb Doukkali University El Jadida, Morocco (e-mail: baddi.y@ucd.ac.ma).

In the following section, we present a set of related works that have proposed a solution for the connected objects classification issue. After, we give an overview of the different feature selection methods. Then in section four, we show the proposed methodology for connected devices classification. Afterward, we present our experimental results and analyze them. Finally, in section 6 a conclusion.

II. Related Works

In recent years, connected devices classification and type identification represent a research subject which occupies a great interest due to its importance for better management of connected devices in a network and contributes to reinforcing devices security, but in the case of using these techniques by attackers, it will be a tool for vulnerable devices detection which presents a major security risk.

Until now, various solutions and methods [5] [6] [7] [8] [9] [10] [11] for the connected device's recognition from different connected devices data like network traffic have been proposed and analyzed in many research works. To identify and classify connected devices, divers methods have been applied such as statistical classification, Random Forest algorithm, Convolutional Neural Network, Artificial Neural Network, XGBoost, Decision Tree, SVM, k-Nearest Neighbours, and Gaussian Naïve Bayes. By applying these algorithms, researchers showed different accuracy for result prediction and propose a set of better models that are suitable for the relevant datasets.

Bai & al. [12] present an end-to-end method for automatic IoT device classification based on deep learning algorithms to identify new and unseen devices by analyzing IoT network traffic. For this, they use traffic streams to characterize the semantic categories of devices and propose an LSTM-CNN cascade model to automatically identify the semantic type of device.

A new solution entitled Acquisitional Rule-based Engine (ARE) has been proposed in [13]. This solution can automatically generate rules for IoT devices discovery and annotation in cyberspace without human effort or training data. ARE is mainly based on the relationship between the application data in IoT devices and the corresponding description websites.

SANTOS et al. [14] have been developed a method that identifies IoT devices and the network flows generated by these devices through traffic analysis. Traffic analysis is performed using two related techniques (statistical classification and the Random Forest algorithm) and features were extracted using the Best First Search model. The proposed method was able to identify 99% of devices and classify their network traffic with great precision.

Another approach for precise IoT devices identification has been developed by Meidan & al. [15] based on network traffic characteristics and machine learning. To determine if the traffic is generated by a specific (known) PC, smartphone, or IoT device, they propose a multistep process (ProfilIoT) in which a set of machine learning classifiers are applied to a flow of sessions from a specific device (that is, a specific IP address). Their method allows the classification of IoT devices (including by brand and model) with an accuracy of 99.281%.

A semi-supervised model for IoT devices identification has been proposed by Fan et al. [16]. This model is based on Convolutional Neural Network and multitasks learning to distinguish IoT and non-IoT devices and also to classify specific IoT devices with a few labeled data. The evaluation of this model in a public dataset shows that 5% labeled data is needed and gets an accuracy over 99%.

Salman et al. [17] present a new framework for IoT devices and traffic type identification, and abnormal traffic detection. This framework has four components: features extractor, IoT device identification, traffic-type identification, and intrusion detection. Experiments have shown that the proposed data representation is efficient for the classification task and that the random forest algorithm gives the best results compared to different machine learning algorithms tested with an accuracy of up to 94.5%.

In another work, Miettinen et al. [18] propose a new approach for automatic connected devices identification in IoT networks, which has entitled IOT SENTINEL. It's a framework that aims to extract features from the flow of packets generated at the first time of devices connected to the network which present a signature of IoT devices, then classify them by types/models using a random forest classifier. The results present that IOT SENTINEL is effective for device type identification and give a great accuracy (81.5% for identification of considered devices) and has a minimal performance overhead.

Marchal et al. [19] propose AUDI that represent an autonomous approach for quick and effective device identification in IoT network. It allows the identification of devices in any operation mode and does not require any knowledge of devices types. AUDI is based on periodic communication traffic of IoT devices and unsupervised learning methods. The proposed technique consists of three main steps: Step 1: the capture of periodic flows, their period and stability, step 2: devices fingerprint extraction, and step 3: device-types fingerprint classification.

System Identifier (SysID) [20] represents another system for automated devices classification based on network traffic and using machine learning and genetic algorithm (GA). To develop this system, authors use GA to select relevant features and eliminate noisy features in the aim to increase classification performance, then they deploy multiple machine learning algorithms for devices classification. SysID allows a completely automated devices identification with higher accuracy (over 95%), but in some cases, it is not able to identify devices from the same vendor due to their network behaviors similarity.

On the other hand, Sivanathan et al. [21] have developed a novel framework for IoT devices classification based on various network traffic characteristics and using machine learning. It consists of three main steps: Step 1: traffic collection and synthesis, step 2: traffic characterization, and step 3: devices identification. This multi-stage framework allows the identification of specific IoT devices with an accuracy of over 99%.

III. Feature Selection

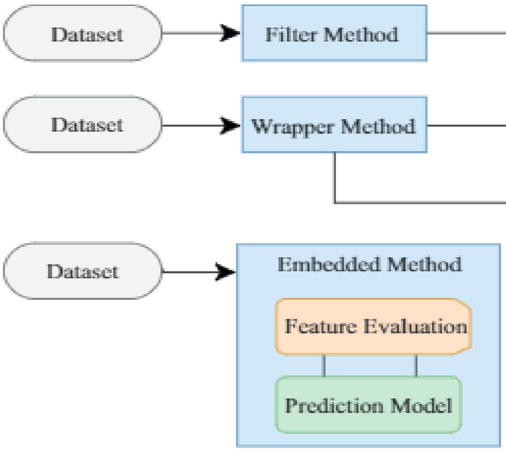

Fig. 1. The main feature selection methods for machine learning [23]

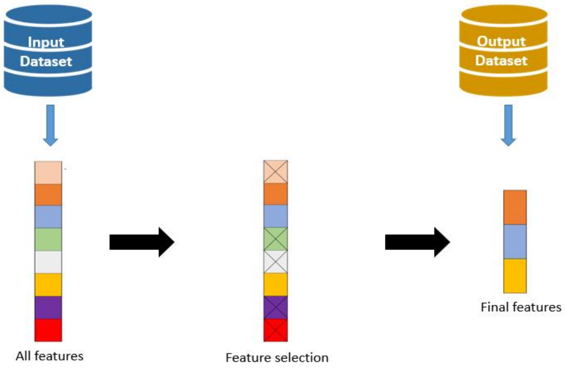

Fig. 2. Feature selection

Feature selection is a process of taking a small subset of features from the original feature set without transformation (thus preserving the interpretation) and validating it concerning the analysis goal. The selection process can be achieved in several ways depending on the goal, the resources at hand, and the desired level of optimization. [22]

Feature selection methods allow the discovery of the most important and representative attributes were to contribute most to your output/targeted variable, identifying and removing irrelevant or less important features from an input set. It must be at the first stage of model designing, can be included in the step of data cleaning, has great importance in any machine learning problem, and is usually used in the classification, regression, and clustering tasks.

We can distinguish three categories of feature selection methods: Filter Methods, Wrapper Methods, and Embedded Methods.

A. Filter Methods:

Based on mathematical and statistical concepts, filter methods select a subset of features according to their relationship with the target. They exist a set of measures for feature filtering that can be classified into information, distance, consistency, similarity, and statistical measures. Filters are based on performance measures without taking into account the used algorithm/model. They are divided into two categories: univariate methods that consider each feature separately and multivariate methods, which evaluate entire feature subsets.

This method was used in different work related to IoT traffic classification and also intrusion detection like the work released by Deka et al. in [24] uses Correlation coefficient, dispersion coefficient, and information entropy in the aim to reorganize their dataset based on features relevancy and non-redundancy. The experimental results show that decreasing the number of features does not deteriorate the accuracy (can give the same accuracy) and can be effective for IoT traffic classification.

B. Wrapper Methods:

Wrappers tend to perform better in selecting features since they take the model hypothesis into account by training and testing in the feature space. This leads to the big disadvantage of wrappers, the computational inefficiency which is more apparent as the feature space grows. Unlike filters, they can detect feature dependencies. Wrappers are separated into 2 categories: Randomised and Deterministic [23]

C. Embedded Methods:

These methods serve to select features that contribute to having the best accuracy of the model by combining the step of feature selection and model building. It contributes better than others methods, but each technique depends on the classifier and does not work with any other classifier. Embedded methods include various techniques like CART, C4.5, and Lasso where random forests represent the well-known technique.

Many works have focused on the importance of using feature selection like [25] [26] [27], which demonstrate that the use of these methods before modeling data allows:

- Achieve efficient data reduction
- Reduces overfitting
- Improve accuracy
- Reduce training time

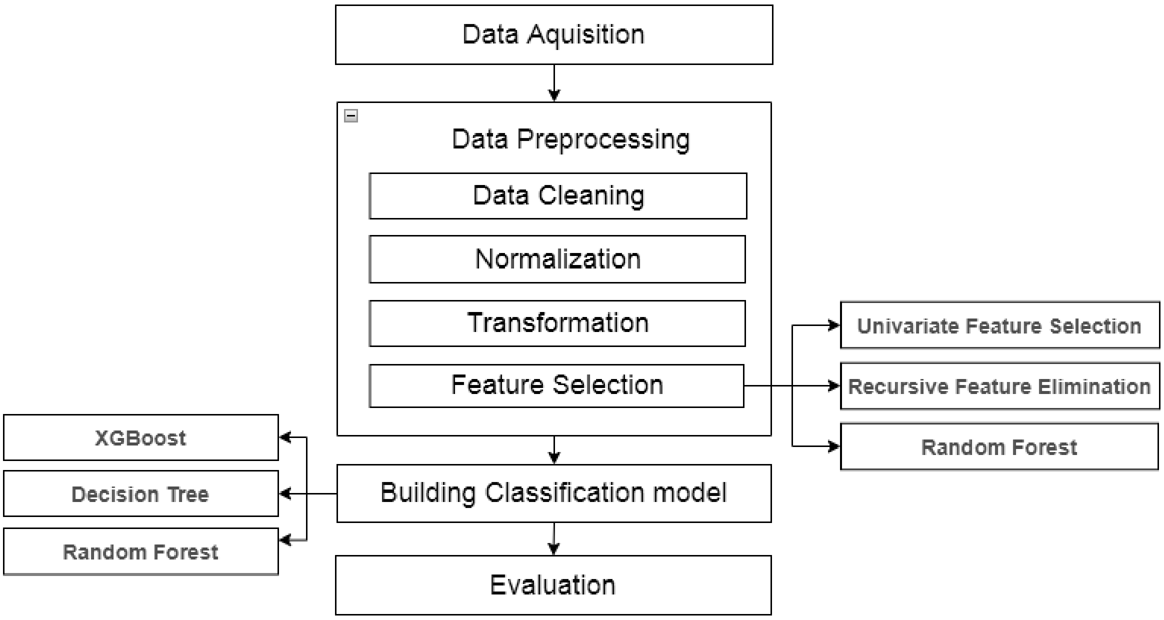

Fig. 3. Overview of our methodology

IV. Methodology

In this section, we present the proposed methodology for connected devices classification that is shown in Figure 3 and consists of four major steps:

- Data Collection and building a dataset represent the first step of our methodology. Our data are collected from the Shodan database (Shodan is a search engine that represents a tool specialized to searching IOT devices based on banners grabbing information to detect and identify the connected devices [28]) which provides a set of information related to discovered devices and services, including Internet Protocol address (IP), port and banner services. In Table I we describe the quantitative properties of our datasets, as well as the feature set of collected data, which is shown in Table II.

TABLE I
Datasets Description

Dataset Features	Parameters of 1st Dataset	Parameters of 2nd Dataset
Total instance	1129	1064
Total training data	847	798
Total testing data	282	266
Total features	18	18
Number of classes	14	14

- The second step is related to data pre-processing, which represents an essential step in machine learning model building. Data pre-processing includes data cleaning, normalization, transformation, feature extraction, selection, etc. [29] Our vision in this step is to prepare our data and select the most representative features of the connected devices, and to have the best results in the connected devices classification task.

TABLE II
Feature set of collected data

Feature	Description
Query	Keyword used to search devices (type of devices)
Type	Device type
IP	Internet Protocol address
ASN	Autonomous System Number
Port	Port number that the service is operating on.
Transport	IP transport protocol used to fetch information
Hostnames	Array of hostnames that have been assigned to the IP address for this device
Domains	Array of the top-level domains for the hostnames of the device
OS	Operating system that powers the device
Organization	name of organization that is assigned the IP space for this device
Product	The name of the product that generated the banner
Version	Version of the product
Data	banner information for the service
SSL	Secure Sockets Layer: If the service uses SSL, such as HTTP then the banner will also contain a property called "ssl"
longitude	longitude for the geolocation of the device
latitude	latitude for the geolocation of the device
City	Name of the city
Country	Name of the country
CPE	relevant Common Platform Enumeration for the product or known vulnerabilities if available
OPTS	Contains experimental and supplemental data for the service

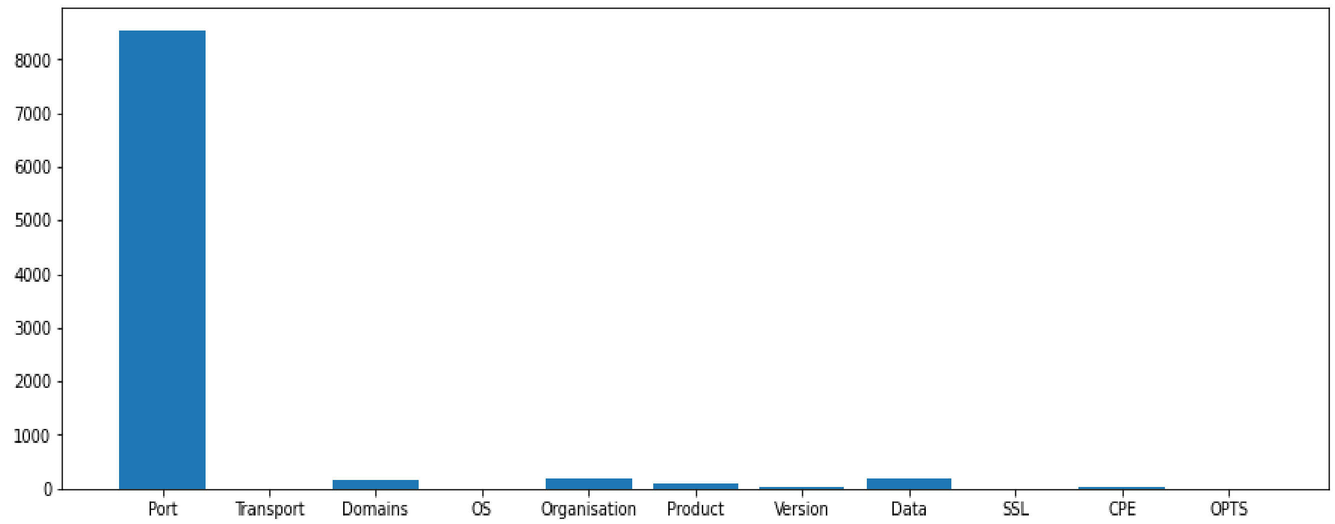

Fig. 4. Univariate feature selection

- The third step is dedicated to building a classification model. In our work, we use three classification models (XGBoost, Decision Tree, and Random Forest) which are mainly used in the classification of IoT devices. The input of the used machine learning algorithms is the feature set selected in the previous step, and the output is the set of connected objects type.
- In the last step, we evaluate the classification models

V. Experimental Results and Discussion

In this section, we study the effect of feature selection methods to achieve high performance of the machine learning algorithms for IoT devices classification. For this purpose, we test three feature selection methods on two datasets collected from shodan API data and by using three machine learning algorithms.

A. Data Preprocessing:

In general, the dataset may contain incomplete, unstructured, irrelevant, inconsistent, and inaccurate data, which could produce bad results as well as lead to poor accuracy and speed of the model. To avoid this problem, we will focus the second step of our process on data pre-processing, which consists in applying various operations to facilitate the learning phase and to have an efficient model. The pre-processing that we release aims to:

- Replace missing values
- Encode Categorical Data
- Remove irrelevant data
- Make data anonymous

B. Feature Selection Methods:

In our experiment, we use the following feature selection methods: Univariate feature selection, Recursive Feature Elimination (RFE), Tree-based feature selection (Random Forest)

- Univariate feature selection elaborates the selection task of the best features based on univariate statistical tests. They are a set of measures to use, in our case the input and output data are categorical for that we use mutual information/ chi-squared test.
- Random forests algorithm is one the most popular machine learning methods used for the classification task. Besides that, it can be used as a feature selection and ranking method. Generally, the random forest-based feature selection approach exposes the impurity-based feature ranking method. Therefore, generally, it requires little feature engineering and parameter tuning. [30]
- Recursive feature elimination (RFE) is based on the idea to select features by repeatedly constructing a model and pruning the worst-performing feature based on coefficients from the current feature sets. This procedure is repeated on the pruned feature set until the intended number of features is reached. The performance of RFE greatly depends on the model used for feature ranking. [30]

C. Feature Selection Results:

In Figure 4, Figure 5, and Figure 6 we illustrate the results of the application of recursive feature elimination, tree-based feature selection method, and univariate feature selection on our first dataset. According to these results, we can conclude that the process of connected devices identification can be done by using just 5 important features for devices representation which are data, product, organization, port, and domains.

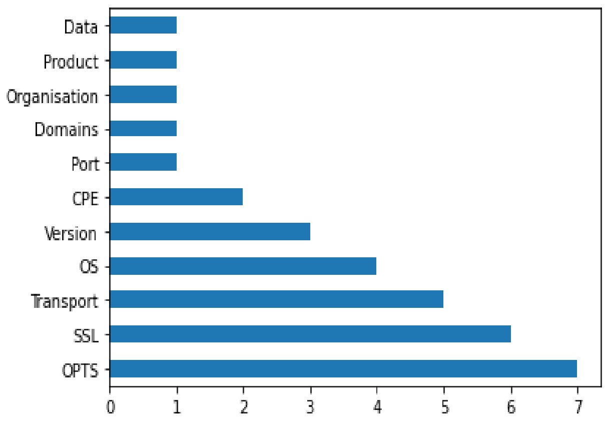

Fig. 5. Recursive Feature Elimination

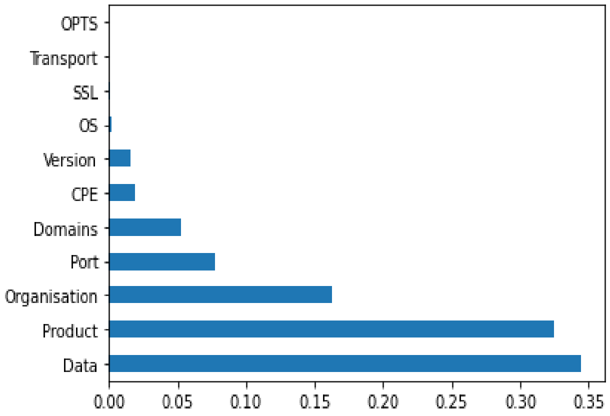

Fig. 6. Tree Based Feature Selection

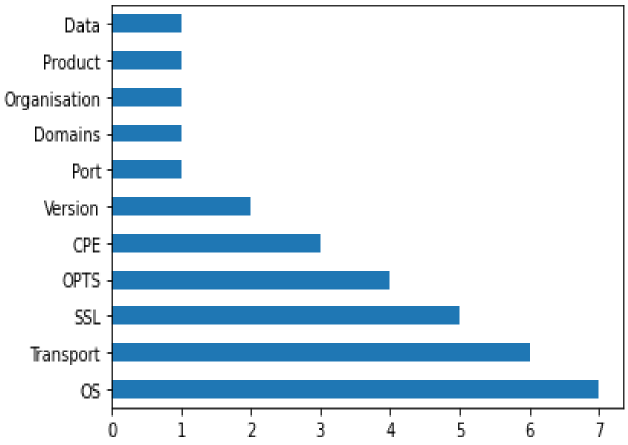

Fig. 7. Recursive Feature Elimination

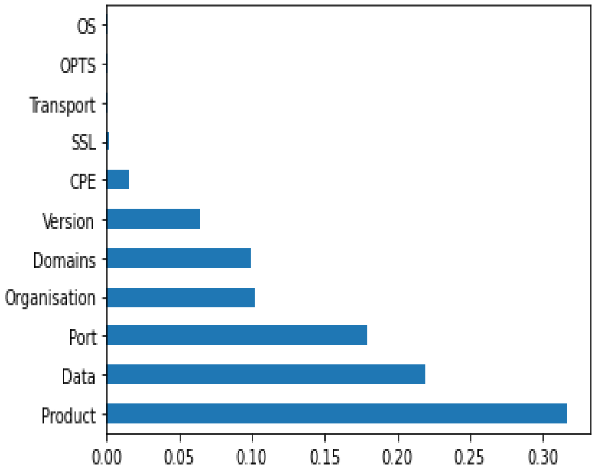

Fig. 8. Tree Based Feature Selection

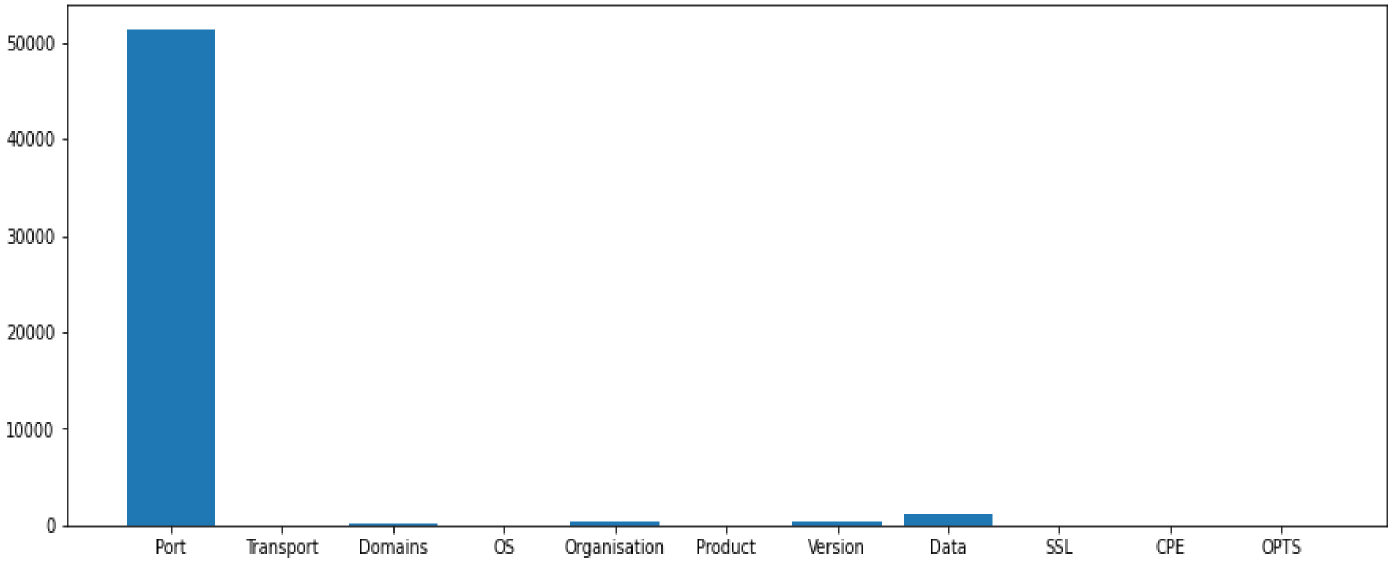

Fig. 9. Univariate feature selection

In Figure 7, Figure 8, and Figure 9, we present the results of applying the three feature selection methods on our second dataset. These results allow us to support the hypothesis that the process of identifying connected devices can be performed using only 5 features important for the representation of connected devices.

D. Evaluation Results:

We use the Hold-out cross-validation approach. This approach serves to split the dataset into two sets:

- Training set: what the model is trained on
- Test set: to see how well that model performs on unseen data

We use the accuracy metric to evaluate our approach:

Accuracy = (TP + TN) ÷ (TP + TN + FP + FN) (1)

TABLE III
ACCURACY OF DIFFERENT FEATURE SELECTION METHODS USING DIFFERENT MACHINE LEARNING ALGORITHMS ON THE 1ST DATASET

	Machine learning Algorithms		
	XGBoost	Random Forest	Decision Tree
Metrics / Feature selection	Accuracy	Accuracy	Accuracy
Univariate	1	0.996	0.996
RFE	1	0.996	0.996
Tree-based	1	1	0.996

TABLE IV
ACCURACY OF DIFFERENT FEATURE SELECTION METHODS USING DIFFERENT MACHINE LEARNING ALGORITHMS ON THE 2ND DATASET

	Machine learning Algorithms		
	XGBoost	Random Forest	Decision Tree
Metrics / Feature selection	Accuracy	Accuracy	Accuracy
Univariate	0.988	0.981	0.977
RFE	0.996	0.996	0.985
Tree-based	0.996	0.996	0.988

We evaluate three classifiers: XGBoost, Decision Tree, and Random Forest. Our particular objective is to obtain the best ranking. In Table III and Table IV, we present the classification accuracy of the machine algorithms tested on the 5 most representative features of our dataset. These results show that:

- Feature selection allows achieving an effective classification of connected devices.
- XGBoost classifier achieves almost perfect classification accuracy for identifying types of devices.
- Tree-based feature selection method achieved the highest classification accuracy with Random Forest and XGBoost

Advantages of the proposed approach:

- This approach is based on a new dataset that includes banner grabbing information
- The high-performance of the proposed approach for the detection and identification of connected objects in the network
- Context of use is multiple, including integration into IoT search engines and reinforcement of network security mechanisms that integrate connected objects

VI. CONCLUSION

The main purpose of this paper was to evaluate the impact of using feature selection with machine learning for efficient connected devices classification and type identification. For this, we examine the combination of three feature selection methods with three machine learning classifiers that are mostly used in the classification task using two datasets based on the banner grabbing information. The evaluation results show that the selection of appropriate features helps to improve the accuracy of connected devices classification using machine learning classifiers. However, further studies need to be conducted on this combination to propose an effective model for connected devices classification based on feature selection and machine learning.

REFERENCES

[1] F. Z. Fagroud, E. Ben Lahmar, and S. El Filali, "Internet of Things: Statistical Study On Research Evolution, " *International Journal of Advances in Electronics and Computer Science*. vol. 6, no. 5, pp. 4-13, 2019.

[2] F. Z. Fagroud, E. Ben Lahmar, H. Toumi, Y. Baddi , and S. El Filali, "RT-RCT: an online tool for real-time retrieval of connected things". *Bulletin of Electrical Engineering and Informatics*. vol. 10, no 5, pp. 2804-2810, 2021.

[3] F. Z. Fagroud, L. Ajallouda, E. Ben Lahmar, H. Toumi, A. Zellou , and S. El Filali, "A Brief Survey on Internet of Things (IoT) ". *In International Conference on Digital Technologies and Applications*. Springer, pp. 335-344, 2021.

[4] S. Mustafizur R., G. Blanc, Z. Zhang, and H. Debar, "Iot devices recognition through network traffic analysis," *in 2018 IEEE International Conference on Big Data (Big Data).* pp. 5187-5192, 2018.

[5] N. Ammar, L. Noirie, and S. Tixeuil, "Identification du type des objets connectés par les informations des protocoles réseaux," *In : Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication*. 2018.

[6] T. JAVED, M. Haseeb, M. Abdullah, and M. Javed, "Using application layer banner data to automatically identify IoT devices," *ACM SIGCOMM Computer Communication Review*. vol. 50, no 3, pp. 23-29, 2020.

[7] K. SINGLA, and J. BOSE, "Iot2vec: Identification of similar iot devices via activity footprints, " *In : 2018 International Conference*

on Advances in Computing, Communications and Informatics (ICACCI). pp. 198-203, 2018.

[8] A. SIVANATHAN, H. H. Gharakheili, and V. Sivaraman, "Inferring iot device types from network behavior using unsupervised clustering," *In : 2019 IEEE 44th Conference on Local Computer Networks (LCN)*. pp. 230-233, 2019.

[9] P. KHANDAIT, N. Hubballi, and B. Mazumdar, " IoTHunter: IoT network traffic classification using device specific keywords, " *IET Networks.*

[10] K. R. Kumar, C. Hemanth, C. A. Kumar, K. M. Sahith, and G. A. Prasanth, "Iot Device Identification Through Network Traffic Analysis".

[11] A. Bremler-Barr, H. Levy, and Z. Yakhini, " Iot or not: Identifying iot devices in a short time scale". *In NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*. IEEE, pp. 1-9, 2020.

[12] L. Bai, L. Yao, S. S., Kanhere, X. Wang, and Z. Yang, "Automatic device classification from network traffic streams of internet of things," *In 2018 IEEE 43rd Conference on Local Computer Networks (LCN).* IEEE, pp. 1-9, 2018.

[13] X. Feng, Q. Li, H. Wang, and L. Sun, "Acquisitional rule-based engine for discovering Internet-of-Things devices," *In 27th {USENIX} Security Symposium ({USENIX} Security 18)*. pp. 327-341, 2018.

[14] S. Matias RP, R. M. Andrade, D. G. Gomes, and A. C. Callado, "An efficient approach for device identification and traffic classification in IoT ecosystems," *in 2018 IEEE Symposium on Computers and Communications (ISCC)*. pp. 00304-00309, 2018.

[15] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, "ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis," *In Proceedings of the symposium on applied computing*. pp. 506-509, 2017.

[16] F. Linna, S. Zhang, Y. Wu, Z. Wang, C. Duan, J. Li, and J. Yang, "An IoT Device Identification Method based on Semi-supervised Learning," *In : 2020 16th International Conference on Network and Service Management (CNSM)*. IEEE, pp. 1-7, 2020.

[17] O. Salman, I. H. Elhajj, A. Chehab, and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection". *Transactions on Emerging Telecommunications Technologies*. pp. e3743, 2019.

[18] M. Miettinen, S. Marchal, I., Hafeez, N. Asokan, A. R. Sadeghi, and S. Tarkoma, "Iot sentinel: Automated device-type identification for security enforcement in iot". *In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, pp. 2177-2184, 2017.

[19] S. Marchal, M. Miettinen, T. D. Nguyen, A. R. Sadeghi, and N. Asokan, "Audi: Toward autonomous iot device-type identification using periodic communication". *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1402-1412, 2019.

[20] A. Aksoy and M. H. Gunes "Automated iot device identification using network traffic". *In ICC 2019-2019 IEEE International Conference on Communications (ICC).* IEEE, pp. 1-7, 2019.

[21] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath,and V. Sivaraman, "Classifying IoT devices in smart environments using network traffic characteristics". *IEEE Transactions on Mobile Computing,* vol. 18, no. 8, pp. 1745-1759, 2018.

[22] J. Alan, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," *In 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*. pp. 1200-1205, 2015.

[23] B. Mohammed, E. Karaarslan, and Y. HOŞCAN, "A hybrid feature-selection approach for finding the digital evidence of web application attacks," *Turkish Journal of Electrical Engineering & Computer Sciences.* vol. 27, no. 6, pp. 4102-4117, 2019.

[24] R. K. Deka, K. P. Kalita, D. K. Bhattacharyya, and D. Boro, “A Smart Feature Reduction Approach to Detect Botnet Attack in IoT”. *In Emerging Technologies for Smart Cities*. Springer, pp. 17-23, 2021.

[25] P. Nimbalkar, and D. Kshirsagar "Feature selection for intrusion detection system in Internet-of-Things (IoT) ". *ICT Express*, vol. 7, no. 2, pp. 177-181, 2021.

[26] N. Hoque, M. Singh, and D.K. Bhattacharyya, "EFS-MI: an ensemble feature selection method for classification". *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 105-118, 2018.

[27] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data". *Computational Statistics & Data Analysis*, vol. 143, pp. 106839, 2020.

[28] F. Z. Fagroud, L. Ajallouda, E. Ben Lahmar, H. Toumi, k. Achtaich , and S. El Filali, "IOT Search Engines: Exploratory Data Analysis," *Procedia Computer Science*, vol. 175, pp. 572-577, 2020.

[29] S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas "Data preprocessing for supervised leaning". *International journal of computer science*, vol. 1, no. 2, pp. 111-117, 2006.

[30] A. Pektaş and T. Acarman, "Effective feature selection for botnet detection based on network flow analysis," *In International Conference Automatics and Informatics*. 2017.

[31] M. Shahbaz, A. Guergachi, A. Noreen and M. Shaheen, “Classification by object recognition in satellite images by using data mining”. *In Proceedings of the World Congress on Engineering*, Vol. 1, pp. 4-6, 2012.

[32] A. A. Syed, Y. Heryadi, and A. W. Lukas, “A Comparison of Machine Learning Classifiers on Laptop Products Classification Task”

[33] Y. Zhong, L. Luo, X. Wang, and J. Yang, “Multi-factor Stock Selection Model Based on Machine Learning”. *Engineering Letters*, vol. 29, no. 1, pp. 177-182, 2021.

[34] L. Zhang, L. Luo, L. Hu, and M. Sun, “An SVM-Based Classification Model for Migration Prediction of Beijing”. *Engineering Letters*, vol. 28, no. 4, pp. 1023-1030, 2020.

[35] Q. Zheng, M. Yang, X. Tian, X. Wang, and D. Wang, “Rethinking the Role of Activation Functions in Deep Convolutional Neural Networks for Image Classification”. *Engineering Letters*, vol. 28, no. 1, pp. 80-92, 2020.

[36] M. Hasan, D. Chowdhury, K. Hasan, and A. S. Md, “Statistical Features Extraction and Performance Analysis of Supervised Classifiers for Non-Intrusive Load Monitoring”. *Engineering Letters*, vol. 27, no. 4, pp. 776-782, 2019.